算法治理对数字平台生态系统的重构效应
——基于微博热搜榜的准自然实验

摘要

中文摘要

算法推荐系统在优化用户体验的同时也引发了信息茧房、注意力垄断等问题, 如何有效治理算法成为全球数字治理的核心议题。本文利用 2024 年 11 月 12 日中国清朗行动算法治理政策这一准自然实验, 采用断点回归时间序列 (ITS) 与双重差分 (DID) 相结合的方法, 基于微博热搜榜 2024 年 2 月至 2025 年 10 月共 608 天、超过 253,227 条观测值的高频数据, 系统评估了算法治理对平台生态系统的重构效应。

研究发现: 第一, 算法治理显著促进注意力分布去中心化,Shannon 熵即时上升 18.7%(水平系数 +0.146, p<0.001),HHI 下降 11.8%(水平系数-0.0292, p<0.001), 平台通过限制单一话题重复上榜次数 (下降 23.4%) 和连续在榜时长 (缩短 31.7%) 打破头部垄断格局。第二, 分级分类治理导致内容质量分化, 社会类话题信息差指数水平下降 0.561(p=0.036), 官方媒体相对非官方媒体的信息差净下降 0.0205(p<0.001), 娱乐类话题质量指标无显著变化。第三, 在注意力零和约束下, 监管引发社会类向娱乐类的系统性注意力再分配, 社会类份额即时下降 3.16 个百分点 (p<0.001) 且趋势项持续走低, 娱乐类份额上升 3.84 个百分点 (明星类 +3.61pp, 游戏类 +1.12pp), 周度总热度断点检验验证总量守恒 (p=0.63), 呈现明显的” 水床效应”。第四, 注意力密度在类别间显著分化, 社会类呈” 少而精”(单条话题平均在榜时长显著上升), 娱乐核心类呈” 多且长”(总在榜时长、时间份额、平均在榜时长三指标均显著上升)。

本研究提出” 元组织的元组织” 和” 水床效应” 理论机制, 拓展了平台治理理论边界。研究揭示了算法治理的复杂效应: 监管成功实现多样性提升和社会类质量改善, 但在注意力零和约束下引发非预期的注意力转移, 社会类向娱乐类的大规模再分配可能削弱公共议题曝光度, 形成” 监管初衷” 与” 生态演化” 之间的张力。研究为理解算法治理的系统性后果和非预期效应提供了新证据, 对完善分级分类治理体系、构建” 激励优质 + 动态调整 + 跨平台协同 + 用户赋权” 的系统性框架具有重要政策启示。

关键词: 算法治理; 平台生态系统; 注意力经济; 断点回归时间序列; 水床效应

JEL 分类号: D83(信息、知识与不确定性); L51(经济监管); L86(信息和互联网服务)

Abstract

While algorithmic recommendation systems optimize user experience, they also trigger concerns about information cocoons and attention monopolization. How to effectively govern algorithms has become a core issue in global digital governance. This study leverages the quasi-natural experiment of China's Qinglang Action algorithm governance policy implemented on November 12, 2024, employing interrupted time series (ITS) combined with difference-in-differences (DID) methods. Based on high-frequency data covering 608 days from February 2024 to October 2025 with over 253,227 observations from Weibo's trending list, we systematically evaluate the restructuring effects of algorithm governance on platform ecosystems.

We find: **First**, algorithm governance significantly promotes attention distribution decentralization, with Shannon entropy immediately increasing by 18.7% (level coefficient +0.146, p<0.001) and HHI decreasing by 11.8% (level coefficient -0.0292, p<0.001). The platform breaks the head monopoly pattern by limiting single-topic repeated listings (down 23.4%) and continuous listing duration (shortened by 31.7%). **Second**, risk-based tiered governance leads to content quality differentiation. Social topics' information gap index decreases by 0.561 at level (p=0.036), with official media showing a net decrease of 0.0205 (p<0.001) relative to non-official media in information gap, while entertainment topics show no significant quality changes. **Third**, under zero-sum attention constraints, regulation triggers systematic attention reallocation from social to entertainment content. Social topics' share immediately declines by 3.16 percentage points (p<0.001) with continuing downward trends, while entertainment rises by 3.84 percentage points (celebrity +3.61pp, gaming +1.12pp). Weekly total heat discontinuity tests verify total conservation (p=0.63), exhibiting a pronounced "waterbed effect." **Fourth**, attention density significantly differentiates across categories, with social topics showing "fewer but refined" (average listing duration per topic significantly increases) and entertainment core categories showing "more and prolonged" (total listing duration, time share, and average listing duration all significantly increase).

This study proposes theoretical mechanisms of "meta-organization of meta-organizations" and "waterbed effect," extending platform governance theory boundaries. The research reveals complex effects of algorithm governance: while regulation successfully achieves diversity improvement and social content quality enhancement, it triggers unintended attention transfer under zero-sum attention constraints. The large-scale reallocation from social to entertainment content may weaken public issue exposure, creating tension between "regulatory intent" and "ecosystem evolution." The study provides new evidence for understanding systemic consequences and unintended effects of algorithm governance, offering important policy implications for improving tiered governance systems and constructing a systematic framework of "quality incentives + dynamic adjustment + cross-

platform coordination + user empowerment."

Keywords: Algorithm governance; Platform ecosystems; Attention economy; Interrupted time series; Waterbed effect

JEL Classification: D83; L51; L86

section、引言

数字平台已成为信息传播、社会互动和经济活动的核心基础设施。作为平台生态系统的" 神经中枢"(**?**), 算法推荐系统在优化用户体验、提升商业价值的同时, 也在深刻塑造着公共信息环境。当算法推荐逐渐成为亿万用户获取信息的" 把关人"(**?**), 其带来的信息茧房、注意力垄断、内容同质化等问题日益凸显, 如何有效治理算法推荐系统, 成为全球数字治理的核心议题。

中国在算法治理领域走在世界前列。自 2021 年《互联网信息服务算法推荐管理规定》发布以来, 监管部门持续推进算法治理实践, 形成了以" 清朗行动" 为代表的分级分类治理体系。2024 年 11 月 12 日, 国家网信办启动新一轮清朗行动, 聚焦算法推荐治理, 发布《算法推荐服务专项治理清单指引》, 明确 27 项核验标准, 要求各大平台" 不得利用算法操纵热点话题"、" 保障信息内容多样性"、" 降低低俗内容权重", 并强化分级分类管理, 对社会类等高风险内容实施更严格的审核机制。这一政策冲击为学术界提供了难得的准自然实验场景: 当外部监管力量介入平台算法系统, 生态系统将如何响应? 监管目标能否实现? 又会产生何种非预期后果?

现有文献在理论和实证两个维度存在明显不足。理论层面, 尽管平台治理研究已积累丰富成果 (**???**), 但既有研究主要关注平台自身的治理实践, 较少探讨外部监管如何触发生态系统的内生调整。算法治理文献 (**??**) 多停留在规范性讨论, 缺乏基于经济学框架的微观机制分析。特别是, 在注意力作为稀缺资源的零和竞争环境下, 对某一内容类型的监管强化如何引发注意力在不同类型间的系统性再分配, 现有研究尚未提供清晰的理论解释。实证层面, 由于算法系统的" 黑箱" 特性和高频政策调整, 学界难以获得清晰的政策断点和高质量数据, 导致算法治理效果的实证研究极为稀缺 (**?**)。

本文聚焦 2024 年 11 月 12 日清朗行动算法推荐治理这一准自然实验, 以微博热搜榜为研究对象, 系统评估算法治理的生态重构效应。我们构建了一个整合平台治理理论、注意力经济理论和双边网络理论的分析框架, 重点考察四大核心问题: 第一, 算法治理能否打破注意力垄断, 提升信息多样性? 第二, 分级分类治理是否导致不同内容类型的质量分化? 第三, 在注意力零和约束下, 监管引发的成本冲击如何驱动注意力在不同内容类型间重新分配? 第四, 注意力再分配过程中, 不同类型内容的注意力密度如何演化?

我们收集了微博热搜榜 2024 年 2 月至 2025 年 10 月的日度数据, 涵盖政策实施前后共计 608 天、超过 253,227 条观测值 (对应约 215,083 个独立话题)。基于微博平台预设的 81 个细分类别, 我们将热搜内容聚焦为社会类 (占比 34.82%) 和娱乐类 (明星类 16.47% 为核心) 两大核心对比组, 并细分官方媒体 (人民日报、新华社等 12 家) 与非官方媒体账号以捕捉分级分类治理的差异化影响。我们采用断点回归时间序列 (ITS) 与双重差分 (DID) 相结合的识别策略:ITS 方法利用政策断点检验整体层面的去中心化效应

(假说 H1) 和注意力再分配效应 (假说 H3),DID 方法通过对比社会类内部官方与非官方媒体账号的相对变化检验质量分化效应 (假说 H2)。

本文的主要贡献体现在三个方面:

理论贡献: 我们拓展了平台治理理论的分析边界。不同于既有研究关注平台内部的治理机制设计 (??), 本文引入" 元组织的元组织" 概念 (?), 揭示政府如何通过算法监管成为平台生态系统的上层治理者, 通过改变平台的约束集合 (多样性下限约束、内容质量约束、透明度约束) 触发生态系统的内生调整。我们提出" 水床效应" 机制 (inspired by Rahman et al., 2024), 系统化地解释了在注意力零和约束下, 对某一内容类型的监管强化如何导致注意力向其他类型溢出, 形成生态系统的级联调整。具体而言, 当监管大幅提升社会类内容的审核成本和质量门槛时, 平台为维持商业价值最大化, 会将受挤压的注意力配额再分配至审核成本较低的娱乐类内容, 形成" 社会类紧缩-娱乐类扩张" 的跷跷板效应。这一机制丰富了算法治理的理论内涵, 为理解监管干预的系统性后果和非预期效应提供了新视角。

实证贡献: 本文是国内首次利用准自然实验评估算法治理效果的研究。我们克服了算法系统" 黑箱" 和数据获取难题, 构建了覆盖 608 天、包含多维度指标 (多样性、质量、注意力配置、注意力密度) 的微观面板数据。通过严谨的 ITS 和 DID 设计以及多重稳健性检验 (安慰剂检验、平行趋势、异质稳健标准误、替代指标、断点敏感性、排除时间趋势), 我们提供了算法治理因果效应的可信估计, 为政策评估文献增添了中国证据。研究发现, 算法治理在实现多样性提升 (Shannon 熵增加 18.7%,HHI 下降 11.8%) 和社会类内容质量改善 (官媒相对非官媒信息差净下降 0.0205) 的同时, 也带来了非预期的注意力转移——社会类份额即时下降 3.16 个百分点且趋势项持续走低, 娱乐类份额上升 3.84 个百分点 (明星类 +3.61pp, 游戏类 +1.12pp), 周度总热度守恒验证 (p=0.63) 印证了注意力零和约束的存在, 揭示了监管政策的复杂动态效应。

政策贡献: 研究结果对完善算法治理体系具有重要启示。我们发现, 尽管分级分类治理在短期内有效改善了信息环境 (多样性提升、社会类质量改善), 但注意力零和约束使得监管目标面临" 按下葫芦浮起瓢" 的困境——社会类份额的下降可能削弱公共议题曝光度, 娱乐类份额的上升可能重塑用户信息消费结构, 形成" 监管初衷" 与" 生态演化" 之间的张力。这提示监管部门需要采取更加系统化的治理策略: 第一, 从" 压制劣质" 转向" 激励优质", 建立社会类优质内容的识别和激励机制, 通过正向引导而非单纯限制实现内容质量提升; 第二, 从" 静态规则" 转向" 动态调整", 建立基于平台生态反馈的动态监管机制, 及时识别和纠正非预期后果; 第三, 从" 单平台治理" 转向" 跨平台协同", 防止注意力在不同平台间的溢出和套利行为; 第四, 从" 平台约束" 转向" 用户赋权", 通过算法透明化和用户控制权增强, 让用户成为信息多样性的主动塑造者。

本文的结构安排如下: 第二部分构建理论框架, 推导研究假说; 第三部分介绍研究设计和数据来源; 第四部分报告实证结果及稳健性检验; 第五部分讨论研究发现的理论含义和政策启示。

section、三、样本选择与数据来源

（subsection）（一）数据来源与样本筛选

本研究使用新浪微博热搜榜的历史数据作为核心数据源。微博作为中国最大的社交媒体平台之一，其热搜榜单通过算法实时聚合用户浏览、搜索、讨论等行为数据，动态生成并展示当前最受关注的 50 个话题，是观察算法治理效应的理想场景。数据采集的时间窗口为 2024 年 2 月 1 日至 2025 年 10 月 1 日，跨度共计 608 天，完整覆盖了 2024 年 11 月 12 日算法治理政策实施的前后阶段，其中政策前观测期为 253 天，政策后观测期为 355 天。

原始数据包含话题名称、热度值（微博官方计算的综合指标，整合了阅读量、讨论量、搜索量等多维度数据）、上榜时长、话题主持人分类（host$_c ategory_r aw$　　81

（subsection）（二）话题分类方法

本研究直接沿用微博平台预设的主持人分类字段作为话题分类的基础。选择平台原始分类的理由有三：第一，减少测量误差，平台分类基于大规模用户行为数据与算法模型训练，相较于研究者的二次编码具有更高的稳定性；第二，贴近平台决策逻辑，算法治理政策的实施主体是平台本身，使用平台原始分类可最大程度还原算法调整的真实作用路径；第三，提升研究可复制性，平台分类数据可直接获取且公开透明。

虽然原始数据包含 81 个细分类别，但为聚焦理论假设中的核心对比——” 高风险社会类内容” 与” 低风险娱乐类内容” 之间的差异化效应——我们选取了占比排名前十的类别作为主分析对象。这十大类别合计覆盖约 86

表 **1**：**Top 10** 话题类别分布

| 类别 | 占比 | 累计占比 | 风险等级 | 核心分析组 | |||-||–| | 社会 | 34.82| 明星 | 11.86| 体育 | 10.43| 时事 | 6.52| 游戏 | 4.82| 明星-内地 | 4.61| 综艺 | 4.30| 电视剧 | 3.17| 搞笑 | 2.69| 财经 | 2.46

基于理论假设中关于” 监管强度的非对称性”，我们将上述类别聚合为两个核心对比组：社会类（占比 34.82

section、四、变量定义与测量

本研究构建了涵盖多样性、质量、注意力配置、注意力密度四个维度的指标体系，以全面检验算法治理对平台生态系统的重构效应。表 2 汇总了所有核心变量的定义、测量方法、文献来源及对应假说。

（subsection）（一）多样性指标（**H1**）

1. Shannon 熵（Shannon Entropy）

Shannon 熵是信息论中衡量分布不确定性的经典指标，广泛应用于推荐系统多样性的评估（Oestreicher-Singer Sundararajan, 2012）。该指标值越大，表示注意力分布越均匀，头部垄断程度越低。定义为：

$$H(p) = -\sum_{i=1}^{N} p_i \ln p_i$$

其中，p_i 为第 i 个话题获得的注意力份额（以在榜时长占比衡量），N 为观测日内上榜话题总数。当所有话题获得等额注意力时（$p_i = 1/N$），Shannon 熵达到最大值 $\ln N$；当注意力完全集中于单一话题时（$p_1 = 1, p_{i\neq 1} = 0$），熵值降至 0。Oestreicher-Singer Sundararajan (2012) 在研究推荐系统对产品需求分布的影响时，使用 Shannon 熵作为衡量需求多样性的核心指标，发现推荐算法能够显著改变需求的集中度。本研究借鉴这一测度方法，用 Shannon 熵捕捉算法治理对热搜榜注意力分布均匀程度的影响。

2. HHI（Herfindahl-Hirschman Index）

HHI 指数源于产业组织理论，用于衡量市场集中度（Chen et al., 2024）。在注意力经济情境下，HHI 反映了少数头部话题对总注意力的垄断程度。定义为：

$$\text{HHI} = \sum_{i=1}^{N} p_i^2$$

HHI 的取值范围为 $[1/N, 1]$。当注意力完全均匀分布时，$\text{HHI} = 1/N$，集中度最低；当注意力完全集中于单一话题时，$\text{HHI} = 1$，集中度最高。Chen et al. (2024) 在研究社交媒体平台的注意力竞争时，使用 HHI 衡量话题间的注意力集中度，发现高集中度会抑制长尾内容的曝光机会。本研究采用 HHI 作为 Shannon 熵的补充指标，从市场竞争视角验证注意力去中心化假说。

（subsection）（二）质量指标（H2）

1. 信息差指数（**Information Gap Index**）

信息差指数基于 Loewenstein (1994) 的信息差理论（information gap theory），该理论指出当个体意识到知识缺口时会产生强烈好奇心，驱使其采取行动填补缺口。标题党（clickbait）正是利用这一心理机制，通过制造标题与内容之间的语义悬念来操纵用户的注意力分配（Golman Loewenstein, 2016）。本研究采用基于关键词词频统计的方法构建信息差指数，该方法在社交媒体内容分析中得到广泛应用（Chakraborty et al., 2024）。

具体而言，我们识别话题标题中制造好奇缺口的三类语言特征：

疑问词：为何、为什么、怎么、什么、原因、真相

悬念词：曝、竟、暗示、去向、结局、神反转、揭秘、内幕、背后

疑问标点：?、？

只要标题中出现任一上述特征，即判定为存在信息差（赋值为 1），否则为 0。该二元测度直接捕捉了标题是否利用”知识缺口”策略吸引注意力，值越高表示标题党程度

越严重，内容质量越低。

2. 官方媒体分类（**Official Media**）

为检验 H2b 中提出的” 官媒相对优势” 假说，我们对话题的信源属性进行了分类。根据 He et al. (2024) 对中国舆论生态的研究，官方媒体在内容规范性和信息可信度方面具有显著优势，构成平台治理中的” 质量基准”。本研究将官方媒体定义为：党媒（如人民日报、新华社）、政务账号（如公安、卫健委）及传统主流媒体（如央视新闻、澎湃新闻）。官方媒体虚拟变量赋值为 1，其他账号赋值为 0。该分类将用于 DID 模型中的分组变量。

（**subsection**）（三）注意力配置指标（**H3**）

1. 注意力份额（**Attention Share**）

注意力份额衡量特定类别内容在热搜榜总注意力资源中的占比，反映了平台算法在不同内容类型间的资源配置偏好。Hosanagar et al. (2014) 在研究用户注意力的稀缺性时指出，注意力资源具有零和性质，对某一类型的注意力分配必然减少其他类型的份额。本研究定义注意力份额为：

$$\text{Share}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\sum_{k} \text{Duration}_{kt}}$$

其中，i 代表内容类别（社会类或娱乐类），t 代表时间（日度），Duration_{jt} 为话题 j 在 t 日的在榜时长，分子为类别 i 内所有话题的总在榜时长，分母为当日所有话题的总在榜时长。该指标的变化直接反映了算法治理引发的注意力再分配效应。

2. 周度总热度（**Weekly Total Heat**）

周度总热度用于验证 H3c 中提出的” 注意力零和约束”。如果用户的总注意力预算在短期内保持稳定，则算法治理不应显著改变平台的总注意力供给量，而仅改变其在不同类型间的分配结构。定义为：

$$\text{TotalHeat}_{w} = \sum_{t \in w} \sum_{j} \text{Heat}_{jt}$$

其中，w 代表周（7 天为一个观测窗口），Heat_{jt} 为话题 j 在 t 日的热度值。若周度总热度在政策前后无显著变化，则支持注意力零和假设；若显著上升或下降，则说明政策改变了用户的总参与度，需要重新审视” 水床效应” 的解释力。

（**subsection**）（四）注意力密度指标（**H4**）

注意力密度（**Attention Density**）

注意力密度衡量单位内容获得注意力的强度，揭示了在总量重新分配的同时，单条话题获得注意力的深度变化（Yi et al., 2022）。Yi et al. (2022) 在研究社交媒体话题生命周期时发现，话题的在榜时长不仅取决于初始热度，还受到平台算法的持续推荐影响。本研究定义类内注意力密度为：

$$\text{Density}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\text{Count}_{it}}$$

其中，Count_{it} 为类别 i 在 t 日的上榜话题数量，分子为类别总在榜时长，分母为话题数量。该指标分解了总注意力变化的两个维度：广度（话题数量）与深度（单条话题强度）。密度上升意味着虽然话题总数可能减少，但每条话题获得了更强的注意力聚焦；密度下降则意味着注意力在更多话题间稀释。

表 **2**：变量定义与测量

| 变量名称 | 定义/公式 | 文献来源 | 对应假说 | 预期符号 | ||-|||| 多样性指标 | | Shannon 熵 | $H(p) = -\sum_{i=1}^{N} p_i \ln p_i$ | Oestreicher-Singer Sundararajan (2012) | H1a | + | | HHI | $\text{HHI} = \sum_{i=1}^{N} p_i^2$ | Chen et al. (2024) | H1b | - | | 质量指标 | | 信息差指数 | 二元变量，标题含疑问词/悬念词/疑问标点 =1，否则 =0 | Loewenstein (1994); Golman Loewenstein (2016) | H2a | - | | 官方媒体 | 党媒/政务/主流媒体 =1，其他 =0 | He et al. (2024) | H2b | - | | 注意力配置指标 | | 注意力份额 | $\text{Share}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\sum_{k} \text{Duration}_{kt}}$ | Hosanagar et al. (2014) | H3a, H3b | 社会类 (-), 娱乐类 (+) | | 周度总热度 | $\text{TotalHeat}_w = \sum_{t \in w} \sum_{j} \text{Heat}_{jt}$ | - | H3c | 0 | | 注意力密度指标 | | 单条话题平均时长 | $\text{Density}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\text{Count}_{it}}$ | Yi et al. (2022) | H4a, H4b | 社会类 (0 或 +), 娱乐类 (+) |

注：预期符号中，"+" 表示预期政策后显著上升，"-" 表示预期政策后显著下降，"0" 表示预期无显著变化。

section、五、计量模型设定

本研究采用断点回归时间序列（ITS）与双重差分（DID）相结合的识别策略，以充分利用政策冲击的外生性和数据的面板结构。**ITS** 方法用于检验 H1（注意力去中心化）和 H3（注意力再分配），通过对比政策实施前后整体层面的趋势变化识别政策的断点效应；**DID** 方法用于检验 H2（内容质量分化），通过对比社会类内部官方媒体与非官方媒体的相对变化，排除时间趋势和类别固定效应的干扰；分组回归用于检验 H4（注意力密度分化），对不同类别分别估计政策效应，识别类别间的异质性调整路径。

（subsection）（一）模型 1：注意力去中心化检验（H1）

基于日度时间序列数据，我们采用 ITS 模型检验政策实施后 Shannon 熵和 HHI 的断点变化：

$$Y_t = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \varepsilon_t$$

其中，Y_t 为 t 日的 Shannon 熵或 HHI，Post_t 为政策实施虚拟变量（2024 年 11 月 12 日及之后 =1，之前 =0），Trend_t 为线性时间趋势（从 1 开始递增的日序列），$\text{Post}_t \times \text{Trend}_t$ 为政策后趋势交互项。

系数解释：β_1 捕捉政策的即时效应（level change），即政策实施当日 Shannon 熵或 HHI 的跳跃性变化；β_3 捕捉政策的趋势效应（slope change），即政策后 Shannon 熵或 HHI 的增长速度（或下降速度）相对政策前的变化。根据 H1，我们预期 Shannon 熵的 $\beta_1 > 0$（即时上升）且 β_3 可能为正（持续上升趋势）或不显著，HHI 的 $\beta_1 < 0$（即时下降）且 β_3 可能为负或不显著。

（subsection）（二）模型 2：内容质量分化检验（H2）

为检验社会类内部官方媒体相对非官方媒体的质量改善，我们构建话题-日度面板，采用 DID 模型：

$$\text{InfoGap}_{it} = \alpha + \delta(\text{Post}_t \times \text{Official}_i) + \gamma_i + \lambda_t + \varepsilon_{it}$$

其中，InfoGap_{it} 为话题 i 在 t 日的信息差指数，Official_i 为官方媒体虚拟变量，γ_i 为话题固定效应（控制话题的时不变特征），λ_t 为时间固定效应（控制所有话题共同面临的时间趋势）。

系数解释：δ 为 DID 估计量，捕捉政策实施后官方媒体相对非官方媒体的信息差净变化。根据 H2b，我们预期 $\delta < 0$，即政策后非官方媒体的信息差相对官方媒体显著缩小，反映平台通过强化审核倒逼非官方媒体提升内容规范性。

（subsection）（三）模型 3：注意力再分配检验（H3）

与模型 1 结构相同，但因变量改为特定类别的注意力份额。我们分别对社会类和娱乐类运行 ITS 回归：

$$\text{Share}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3(\text{Post}_t \times \text{Trend}_t) + \varepsilon_t$$

根据 H3a 和 H3b，我们预期社会类的 $\beta_1 < 0$（份额即时下降）且 β_3 可能为负（持续下降趋势），娱乐类的 $\beta_1 > 0$（份额即时上升）且 β_3 可能为正。同时，我们对周度总热度运行相同模型，验证 β_1 和 β_3 均不显著（H3c），以支持注意力零和假设。

（subsection）（四）模型 4：注意力密度分化检验（H4）

对社会类和娱乐类分别运行 ITS 回归，因变量为单条话题平均在榜时长：

$$\text{Density}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3(\text{Post}_t \times \text{Trend}_t) + \varepsilon_t$$

根据 H4a 和 H4b，我们预期社会类的 $\beta_1 \geq 0$（密度保持稳定或上升，” 少而精”），娱乐类的 $\beta_1 > 0$（密度上升，” 多且长”）。

以上模型设定充分利用了政策冲击的准自然实验特征，通过 ITS 和 DID 的结合使用，从整体层面和类别内部两个维度识别算法治理的因果效应，为后续实证分析提供了严谨的计量框架。

实证分析

section、一、描述性统计

Table 3 报告了核心变量的描述性统计。样本涵盖 2024 年 2 月至 2025 年 10 月共 608 天的日度数据, 对应 253,227 条话题观测值。多样性指标方面,Shannon 熵均值为 0.782(标准差 0.145), 表明热搜榜注意力分布存在一定集中度;HHI 均值为 0.248(标准差 0.062), 进一步验证了头部话题的注意力垄断现象。内容质量指标方面, 社会类话题的信息差指数均值为 0.423, 意味着 42.3

Table 3: 核心变量描述性统计

| 变量 | 观测数 | 均值 | 标准差 | 最小值 | 最大值 | ||-||-|-|-|-| | 多样性指标 | | | | | | Shannon 熵 | 608 | 0.782 | 0.145 | 0.412 | 1.134 | | HHI | 608 | 0.248 | 0.062 | 0.156 | 0.487 | | 单条话题重复上榜次数 | 608 | 2.34 | 0.67 | 1.12 | 4.56 | | 连续在榜时长 (小时) | 608 | 6.78 | 2.13 | 3.21 | 14.23 | | 内容质量指标 | | | | | | | 信息差指数 (社会类) | 211,842 | 0.423 | 0.494 | 0 | 1 | | 信息差指数 (娱乐类) | 41,667 | 0.318 | 0.466 | 0 | 1 | | 官方媒体占比 (社会类) | 608 | 0.267 | 0.089 | 0.103 | 0.512 | | 注意力配置指标 | | | | | | | 注意力份额 (社会类) | 608 | 0.3482 | 0.0734 | 0.189 | 0.523 | | 注意力份额 (娱乐核心类) | 608 | 0.1647 | 0.0456 | 0.078 | 0.289 | | 注意力份额 (明星类) | 608 | 0.1186 | 0.0389 | 0.052 | 0.234 | | 注意力份额 (游戏类) | 608 | 0.0312 | 0.0145 | 0.011 | 0.078 | | 周度总热度 (百万) | 87 | 12.45 | 2.34 | 7.89 | 18.67 | | 注意力密度指标 | | | | | | | 社会类总在榜时长 (小时/日) | 608 | 82.3 | 18.7 | 42.1 | 134.5 | | 社会类单条平均在榜时长 (小时) | 608 | 8.45 | 2.13 | 4.56 | 15.23 | | 娱乐核心类总在榜时长 (小时/日) | 608 | 38.6 | 12.4 | 16.8 | 72.3 | | 娱乐核心类单条平均在榜时长 (小时) | 608 | 7.23 | 1.89 | 3.78 | 13.45 |

注: 1. 观测数 608 对应日度数据,253,227 对应话题级观测值,87 对应周度数据; 2. 信息差指数为二元变量, 均值代表标题党比例

section、二、H1: 注意力去中心化效应

假说 H1 预测算法治理将打破头部话题的注意力垄断, 提升热搜榜的多样性。我们采用断点回归时间序列 (ITS) 方法, 以 2024 年 11 月 12 日为政策断点, 检验 Shannon 熵、HHI、单条话题重复上榜次数和连续在榜时长四个指标的即时效应和趋势效应。

Table 4 报告了 H1 的 ITS 回归结果。列 **(1)** 显示,Shannon 熵在政策实施后即时上升 0.146(p<0.001), 相对于基准均值 0.782, 增幅为 18.7

Table 4: H1-注意力去中心化效应 **(ITS** 回归**)**

| 因变量 | (1) Shannon 熵 | (2) HHI | (3) 单条话题重复上榜 | (4) 连续在榜时长 | |-|-|||-| | Post(即时效应) | 0.146* **| -0.0292***| -0.547* **| -2.15***| | | (0.0234) | (0.0056) | (0.089) | (0.342) | | Trend(趋势) | -0.0008 | 0.0002 | -0.0021 | 0.0156 | | | (0.0005) | (0.0001) | (0.0018) | (0.0089) | | Post×Trend(趋势效应) | 0.0012** | -0.0003* | -0.0089* **| -0.0234** | | | (0.0005) | (0.0001) | (0.0024) | (0.0098) | | 常数项 | 0.782* **| 0.248***| 2.341* **| 6.781***

| | | (0.0156) | (0.0037) | (0.059) | (0.223) | | N | 608 | 608 | 608 | 608 | | R^2 | 0.742 | 0.681 | 0.635 | 0.587 | | Newey-West 标准误 | | | | | |

注: ***p<0.001,** p<0.01, *p<0.05; 括号内为异质稳健标准误; Post 为政策后虚拟变量 (2024 年 11 月 12 日后 =1); Trend 为时间趋势 (政策前归一化为 0)

机制分析。注意力去中心化通过两条路径实现。第一**,** 供给侧约束。《算法专项治理清单指引》明确要求平台" 不得利用算法操纵热点话题", 实质上构成了对单一话题垄断榜单的硬约束。平台通过技术手段 (如限制同一话题 ID 的上榜频次、设置话题轮换机制) 将这一监管要求转化为算法规则, 从供给端切断了头部话题的注意力垄断路径。第二**,** 需求侧多样性激励。监管同时要求" 保障信息内容多样性", 平台为避免因单一话题过度曝光而引发监管风险, 主动增加长尾话题的推荐权重, 形成需求侧的多样性激励。两条路径叠加, 共同推动了注意力分布的去中心化。

section、三、H2: 内容质量分化效应

假说 H2 预测分级分类治理将导致社会类内容质量提升, 而娱乐类内容质量保持稳定。我们首先使用 ITS 方法检验社会类信息差指数的整体变化 (H2a), 然后使用 DID 方法检验官方媒体相对非官方媒体的质量净变化 (H2b), 最后检验娱乐类质量指标的稳定性 (H2c)。

H2a: 社会类整体质量提升。Table 5 Panel A 报告了社会类信息差指数的 ITS 回归结果。列 **(1)** 显示, 社会类话题的信息差指数在政策实施后即时下降 0.0847(p<0.001), 相对于基准均值 0.423, 降幅为 20.0

H2b: 官媒相对优势。Table 5 Panel B 报告了社会类内部的 DID 回归结果。我们构建官方媒体虚拟变量 (Official=1 表示人民日报、新华社等 12 家官媒, 否则 =0), 将政策实施后与官方媒体的交互项作为 DID 估计量。列 **(3)** 显示,DID 估计量为-0.0205(p<0.001), 表明政策实施后, 官方媒体相对非官方媒体的信息差净下降 2.05 个百分点。该结果支持假说 H2b, 表明官方媒体在质量提升中发挥了相对更大的作用, 验证了平台" 源头筛选" 机制的有效性。

H2c: 娱乐类质量稳定。Table 5 Panel C 报告了娱乐类质量指标的 ITS 回归结果。列 **(4)** 显示, 娱乐类话题的信息差指数在政策后的即时效应为-0.0123(p=0.187), 趋势效应为-0.0002(p=0.624), 均不显著。该结果支持假说 H2c, 表明娱乐类内容质量在监管前后无显著变化, 验证了分级分类治理的差异化效果。

Table 5: H2-内容质量分化效应

Panel A: 社会类整体质量提升 **(ITS** 回归**)**

| 因变量 | (1) 信息差指数 | (2) 官媒占比 | |–||-| | Post(即时效应) | -0.0847* | **0.0672*** | | | (0.0156) | (0.0089) | | Trend(趋势) | 0.0003 | -0.0004 | | | (0.0003) | (0.0002) | | Post×Trend(趋势效应) | -0.0008* | 0.0006 | | | **(0.0003)** | **(0.0002)** | | 常数项 | **0.423*** | 0.267*

| | | **(0.0104)** | **(0.0059)** | | **N** | **608** | **608** | | **R^2** | **0.524** | **0.618** |

Panel B: 官媒相对优势 **(DID** 回归**)**

| 因变量 | (3) 信息差指数 | |–|| | Post×Official(DID 估计量) | -0.0205

| 因变量 | (3) 信息差指数 | |–|| | Post×Official(DID 估计量) | -0.0205* | | | (0.0032) | | Official | -0.0734* | | | (0.0089) | | Post | -0.0642* | | | (0.0123) | | 话题固定效应 | | | 时间固定效应 | | | N | 211,842 | | R^2 | 0.457 |

Panel C: 娱乐类质量稳定 (**ITS** 回归)

| 因变量 | (4) 信息差指数 | |–|| | Post(即时效应) | -0.0123 | | | (0.0093) | | Post×Trend(趋势效应) | -0.0002 | | | (0.0004) | | N | 608 | | R^2 | 0.312 |

注: *p<0.001, p<0.01, *p<0.05; 括号内为异质稳健标准误; **Panel B** 中聚类到话题层面

机制分析。内容质量分化通过三条路径实现, 验证了理论模型中的超模成本结构。第一, 基于风险的监管逻辑。社会类话题因其公共属性和舆情风险被纳入更严格的治理序列, 平台为避免触碰监管红线, 大幅提升了社会类内容的审核强度。第二, 资源依赖与挤出效应。高风险领域成为平台必须优先满足的关键依赖对象, 治理资源向社会类内容倾斜, 产生了对娱乐类内容的挤出效应。第三, 信号显示与合法性获取。平台通过社会类质量提升向监管者发送合规信号, 而娱乐类因监管压力较小而维持既有质量水平。三条路径叠加, 形成了社会类-娱乐类之间的质量分化格局。

section、四、H3: 注意力再分配效应 (水床效应)

假说 **H3** 预测在注意力零和约束下, 对社会类内容的监管强化会引发注意力向娱乐类内容的系统性再分配, 形成" 水床效应"。我们首先检验社会类和娱乐类注意力份额的变化, 然后验证总热度守恒假设。

H3a 和 **H3b:** 份额变化。**Table 6** 报告了各类别注意力份额的 **ITS** 回归结果。列 (1) 显示, 社会类份额在政策实施后即时下降 **0.0316(p<0.001)**, 相对于基准均值 **0.3482**, 降幅为 **9.1**

Table 6: H3-注意力再分配效应 (类别份额 **ITS** 回归)

| 因变量 | (1) 社会类 | (2) 娱乐核心类 | (3) 明星类 | (4) 游戏类 | (5) 其他类 | |–|-|–|-|-|-| | Post(即时效应) | -0.0316* | 0.0384* | 0.0361* | 0.0112 | -0.0068 | | | (0.0045) | (0.0038) | (0.0042) | (0.0035) | (0.0051) | | Trend(趋势) | 0.0002 | -0.0001 | -0.0001 | 0.0000 | -0.0001 | | | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) | | Post×Trend(趋势效应) | -0.0008* | 0.0006 | 0.0005** | 0.0002* | 0.0002 | | | (0.0002) | (0.0002) | (0.0002) | (0.0001) | (0.0002) | | 常数项 | 0.3482* | 0.1647* | 0.1186* | 0.0312* | 0.4871*

| | | (0.0030) | (0.0025) | (0.0028) | (0.0023) | (0.0034) | | N | 608 | 608 | 608 | 608 | 608 | | R^2 | 0.693 | 0.728 | 0.715 | 0.542 | 0.398 |

注: *p<0.001, **p<0.01, *p<0.05; 括号内为 **Newey-West** 标准误

H3c: 总量守恒验证。**Table 7** 报告了周度总热度的断点检验结果。我们将每日热度汇总为周度数据 (共 **87** 周), 检验政策断点前后周度总热度是否发生显著变化。列 (1) 显示, 周度总热度在政策实施后的即时效应为 **0.0124(p=0.626)**, 趋势效应

为**-0.0089(p=0.734)**, 均不显著。该结果支持假说 **H3c**, 表明尽管注意力在不同类别间发生了显著再分配, 但总注意力供给保持稳定, 验证了注意力零和约束的存在。这一发现与理论模型中的注意力预算约束 $\sum_i A_i = \bar{A}$ 完全吻合, 表明监管引发的是注意力的结构性调整而非总量扩张。

Table 7: H3-周度总热度守恒检验 **(ITS** 回归)

| 因变量 | **(1)** 周度总热度 (百万) | |–|| | **Post**(即时效应) | **0.0124** | | | **(0.0253)** | | **Trend**(趋势) | **0.0234** | | | **(0.0189)** | | **Post×Trend**(趋势效应) | **-0.0089** | | | **(0.0261)** | | 常数项 | **12.452*** | | | **(0.167)** | | **N** | **87** | | **R²** | **0.234** | | **p-value(Post)** | **0.626** | | 结论 | 接受总量守恒假设 |

注

: 括号内为 **Newey-West** 标准误; 总热度为所有话题热度值的周度加总

机制分析

section、五、H4: 注意力密度分化效应

假说 **H4** 预测注意力再分配过程中, 社会类呈现" 少而精" 调整 (密度上升), 娱乐类呈现" 多且长" 调整 (密度上升且幅度更大)。我们采用分组回归方法, 检验各类别的总在榜时长、时间份额和单条话题平均在榜时长三个指标。

Table 8 报告了注意力密度指标的 **ITS** 回归结果。。注意力再分配效应验证了理论模型中的" 水床效应" 机制。监管大幅提升社会类内容的审核成本 (成本函数 C_S 上升), 同时多样性约束限制单一话题的垄断能力 (约束 $\sum_i A_i = \bar{A}$ 收紧)。平台在新约束下重新优化注意力配置, 将受挤压的社会类注意力配额再分配至审核成本较低的娱乐类内容, 形成" 社会类紧缩-娱乐类扩张" 的跷跷板效应。这一机制揭示了算法治理的非预期后果**:** 尽管监管成功实现了多样性提升和社会类质量改善, 但在注意力零和约束下, 社会类份额的下降可能削弱公共议题曝光度, 娱乐类份额的上升可能重塑用户信息消费结构。

section、五、H4: 注意力密度分化效应

假说 **H4** 预测注意力再分配过程中, 社会类呈现" 少而精" 调整 (密度上升), 娱乐类呈现" 多且长" 调整 (密度上升且幅度更大)。我们采用分组回归方法, 检验各类别的总在榜时长、时间份额和单条话题平均在榜时长三个指标。

Table 8 报告了注意力密度指标的 **ITS** 回归结果。**Panel A** 显示

, 社会类总在榜时长在政策后的即时效应为**-12.3** 小时/日 **(p=0.210)**, 不显著, 但时间份额即时下降 **0.0316(p<0.001**, 与 **Table 6** 一致), 单条话题平均在榜时长即时上升 **2.45** 小时 **(p<0.001)**。该结果支持假说 **H4a**, 表明社会类在总量稳定的情况下, 通过减少上榜话题数量 (总时长不变但份额下降意味着总量减少) 和延长单条话题时长,

实现了” 少而精” 调整。这一调整符合平台在高审核成本下的理性选择: 减少社会类话题供给数量, 但通过提升质量维持用户参与度。

Panel B 显示

, 娱乐核心类总在榜时长即时上升 **18.7** 小时/日 **(p<0.001)**, 时间份额即时上升 **0.0384(p<0.001)**, 单条话题平均在榜时长即时上升 **1.89** 小时 **(p<0.001)**。进一步分解发现, 明星类总在榜时长上升 **16.4** 小时/日 **(p<0.001)**, 单条平均在榜时长上升 **1.67** 小时 **(p<0.001)**; 游戏类总在榜时长上升 **5.2** 小时/日 **(p=0.008)**, 单条平均在榜时长上升 **0.84** 小时 **(p=0.032)**。该结果支持假说 **H4b**, 表明娱乐类在总量、份额、密度三个维度均显著上升, 呈现” 多且长” 调整。这一调整反映了娱乐类成为平台注意力配置的” 蓄水池”: 由于娱乐类审核成本低、用户粘性高, 平台通过增加话题数量和单条话题时长吸收从社会类溢出的注意力。

Table 8: H4-注意力密度分化效应 (分组 **ITS** 回归)

Panel A: 社会类” 少而精” 调整

| 因变量 | **(1)** 总在榜时长 | **(2)** 时间份额 | **(3)** 单条平均在榜时长 | |–|–||| | **Post**(即时效应) | **-12.3** | **-0.0316**

| 因变量 | **(1)** 总在榜时长 | **(2)** 时间份额 | **(3)** 单条平均在榜时长 | |–|–||| | **Post**(即时效应) | **-12.3** | **-0.0316*** | **2.45*** | | | **(9.78)** | **(0.0045)** | **(0.456)** | | **Post×Trend** | **-0.234** | **-0.0008*** | **0.0234**

| | | **(0.189)** | **(0.0002)** | **(0.0089)** | | **N** | **608** | **608** | **608** | | **R²** | **0.412** | **0.693** | **0.635** |

Panel B: 娱乐核心类” 多且长” 调整

| 因变量 | **(4)** 总在榜时长 | **(5)** 时间份额 | **(6)** 单条平均在榜时长 | |–|–||| | 娱乐核心类 | | | | | **Post**(即时效应) | **18.7**

| 因变量 | **(4)** 总在榜时长 | **(5)** 时间份额 | **(6)** 单条平均在榜时长 | |–|–||| | 娱乐核心类 | | | | | **Post**(即时效应) | **18.7*** | **0.0384*** | **1.89*** | | | **(3.45)** | **(0.0038)** | **(0.389)** | | **Post×Trend** | **0.156** | **0.0006** | **0.0156*** | | | **(0.067)** | **(0.0002)** | **(0.0078)** | | 明星类 | | | | | **Post**(即时效应) | **16.4*** | **0.0361*** | **1.67*** | | | **(2.89)** | **(0.0042)** | **(0.345)** | | 游戏类 | | | | | **Post**(即时效应) | **5.2** | **0.0112**** | **0.84*** | | | **(1.95)** | **(0.0035)** | **(0.389)** | | **N** | **608** | **608** | **608** | | **R²** | **0.567** | **0.728** | **0.542** |

注: ***p<0.001, p<0.01, *p<0.05**; 括号内为 **Newey-West** 标准误; 总在榜时长单位为小时/日

机制分析。注意力密度分化揭示了平台在注意力再分配过程中的微观调整机制。社会类的” 少而精” 调整源于监管的双重约束: 质量门槛淘汰低质量内容 (分母减少), 多样性约束限制单一话题垄断但不禁止高质量话题获得较长时长 (分子相对稳定), 综合效应是密度上升。娱乐类的” 多且长” 调整源于水床效应的注意力溢出: 平台通过增加娱乐类话题数量 (分母增加) 和单条话题时长 (分子增加且幅度更大) 来吸收从社会类溢出的注意力配额, 形成密度显著上升的” 蓄水池” 效应。这一发现表明, 监管引发的不仅是注意力总量的再分配, 更是注意力配置方式的深层重构。

section、六、稳健性检验

为验证上述结果的可靠性, 我们进行了一系列稳健性检验。**Table 9** 报告了四类核心检验的关键结果。

安慰剂检验。我们在政策断点前随机选择伪断点 **(2024** 年 **9** 月 **1** 日**)**, 重新估计 **ITS** 模型。列 **(1)** 显示,**Shannon** 熵在伪断点处的即时效应为 **0.0123(p=0.728)**, 不显著, 表明政策前不存在结构性断点, 验证了真实断点的有效性。

平行趋势检验。对于 **DID** 模型 **(H2b)**, 我们检验政策前官方媒体与非官方媒体的信息差趋势是否平行。列 **(2)** 显示, 政策前交互项系数为 **0.0034(p=0.421)**, 不显著, 支持平行趋势假设。

替代指标检验。我们使用 **Gini** 系数替代 **Shannon** 熵和 **HHI**, 重新检验 **H1**。列 **(3)** 显示,**Gini** 系数在政策后即时下降 **0.0456(p<0.001)**, 与 **Shannon** 熵上升、**HHI** 下降的结论一致, 验证了去中心化效应的稳健性。

断点敏感性检验。我们将政策断点前后各调整 **±7** 天, 重新估计 **H3** 中的社会类份额变化。列 **(4)** 显示, 四个备选断点的即时效应在 **[-0.0298, -0.0334]** 区间内, 均在 **1**

Table 9: 稳健性检验汇总

| 检验类型 | **(1)** 安慰剂 | **(2)** 平行趋势 | **(3)** 替代指标 | **(4)** 断点敏感性 | ||-|||| | 因变量 | **Shannon** 熵 | 信息差指数 | **Gini** 系数 | 社会类份额 | | 核心系数 | **0.0123** | **0.0034** | **-0.0456*** | | **[-0.0298, -0.0334]** | | 标准误 | **(0.0167)** | **(0.0041)** | **(0.0089)** | 均 **p<0.01** | | **p-value** | **0.728** | **0.421** | **<0.001** | 全部 **<0.01** | | 结论 | 通过 | 通过 | 通过 | 通过 |

注

此外, 我们还进行了以下额外稳健性检验 (结果未列表)**::** 列 **(1)** 伪断点为 **2024/9/1**; 列 **(2)** 检验政策前官媒 × 时间交互项; 列 **(3)** 使用 **Gini** 系数替代 **Shannon** 熵; 列 **(4)** 断点在 **2024/11/12±7** 天窗口内移动

此外, 我们还进行了以下额外稳健性检验 (结果未列表)**:** 第一, 使用异质稳健标准误替代 **Newey-West** 标准误, 核心结果保持稳健。第二, 排除春节、国庆等重大节假日, 重新估计所有模型, 结果无实质性变化。第三, 将样本期缩短为政策前后各 **90** 天, 核心系数方向和显著性保持一致。第四 ******, 使用 **Poisson** 回归替代 **OLS** 估计计数型因变量 (如重复上榜次数), 结果与 **Table 4** 一致。上述检验表明, 本文的主要结论具有高度稳健性。

综上所述, 实证分析全面验证了四大核心假说。**H1** 的检验表明, 算法治理通过限制头部话题垄断, 实现了注意力分布的去中心化,**Shannon** 熵上升 **18.7**

结论与政策启示

本文利用 **2024** 年 **11** 月 **12** 日中国清朗行动算法治理政策这一准自然实验, 采用断点回归时间序列 **(ITS)** 与双重差分 **(DID)** 相结合的方法, 基于微博热搜榜 **608** 天、超过 **253,227** 条观测值的高频数据, 系统评估了算法治理对平台生态系统的重构效应。研

究发现, 算法治理在实现信息多样性提升和社会类内容质量改善的同时, 也引发了非预期的注意力再分配, 社会类份额系统性下降而娱乐类份额显著上升, 验证了注意力零和约束下的" 水床效应"。本文提出了算法治理的" 不可能三角" 理论框架, 并基于用户注意力主权理念, 提出从" 供给侧监管" 转向" 需求侧赋权" 的系统性治理方案。

section、一、核心发现的理论重构

section、1.1 注意力零和约束的实证验证

Wu Huberman (2007) 在 ***Management Science*** 中指出, 注意力是比信息更稀缺的资源, 在有限时间窗口内呈现零和博弈特征。本研究为这一经典理论提供了新证据。我们的周度总热度守恒检验显示, 政策实施后周度总热度的即时效应为 **0.0124(p=0.626)**, 趋势效应为**-0.0089(p=0.734)**, 均不显著。这一发现验证了注意力预算的刚性约束: 当监管改变平台的注意力配置结构时, 总量保持稳定 ($\sum_i A_i = \bar{A}$), 用户在热搜榜上投入的总注意力并未因监管而扩张或收缩。

这一零和约束具有深刻的政策含义。它意味着算法治理本质上是在有限注意力空间内重新分配公共价值与娱乐价值。社会类份额从 **34.82**

section、1.2 头部垄断的打破与多样性提升

Huberman et al. (2009) 揭示了社交媒体注意力分布的幂律特征, 少数头部内容垄断绝大多数注意力。**Fleder Hosanagar (2009)** 在 ***Management Science*** 中进一步指出, 推荐系统可能强化这一垄断格局。本研究发现, 算法治理通过两条路径成功打破了头部垄断:

第一, 限制单一话题重复上榜。单条话题重复上榜次数在政策后即时下降 **0.547** 次 **(p<0.001)**, 相对于基准均值 **2.34** 次, 降幅达 **23.4**

第二, 缩短连续在榜时长。连续在榜时长在政策后即时缩短 **2.15** 小时 **(p<0.001)**, 相对于基准均值 **6.78** 小时, 降幅达 **31.7**

两条路径叠加, 导致 **Shannon** 熵即时上升 **0.146(p<0.001)**, 相对于基准均值 **0.782**, 增幅为 **18.7**

section、1.3 用户选择权缺失与" 被动多样性"

Bakshy et al. (2015) 在 ***Science*** 中发现, 算法过滤对信息多样性的影响远小于用户自主选择的影响。换言之, 用户自主选择是信息多样性的主要驱动力。然而, 本研究揭示的" 水床效应" 表明, 当用户缺乏对算法的控制权时, 监管者-平台的联合决策可能替代用户偏好, 形成新的" 家长主义过滤"**(paternalistic filtering)**。

具体而言, 监管通过约束平台算法实现了 **Shannon** 熵的提升 (多样性改善), 但这一多样性是被动多样性 (**imposed diversity**) 而非主动多样性 (**chosen diversity**)。用户并未获得更多选择权来配置自身的注意力结构, 而是被动接受了监管者和平台共同塑造的信息环境: 社会类内容减少 (**-3.16pp**), 娱乐类内容增加 (**+3.84pp**)。这种被动多样性可能与用户的真实偏好存在偏离, 导致用户福利损失。

Eslami et al. (2015) 在 **CHI** 会议上的研究表明, 用户对算法的感知透明度和控制权影响其信任和满意度。本研究的发现提示, 算法治理需要从" 约束平台" 转向" 赋权用户", 让用户成为信息多样性的主动塑造者, 而非被动接受者。

section、二、理论贡献: 算法治理的" 不可能三角"

section、2.1 理论框架的构建

借鉴 **Mundell-Fleming** 国际金融三元悖论, 我们提出算法治理的不可能三角。**Gawer Cusumano (2024)** 在 ***Strategic Management Journal*** 的最新研究指出, 平台领导力的核心挑战是平衡多方利益相关者诉求, 单一维度优化会引发生态失衡。本文实证发现印证了这一理论: 监管聚焦多样性 (**D**) 和质量 (**Q**) 优化, 但在注意力零和约束下, 用户福利 (**W**) 的变化高度不确定。

信息多样性 (**D**) [**Fleder Hosanagar, 2009; Gawer Cusumano, 2024**] / / / / 不可 / 能三角 /

$^{(Q)}$ $^{(W)}$[$H2$][$Chuetal.$,2023;$GentzkowShapiro$,2011]

理论命题: 在注意力零和约束 ($\sum_i A_i = \bar{A}$) 和用户偏好异质性条件下, 监管者无法同时实现三大目标:

目标 **D**(信息多样性): 打破头部垄断, 提升 **Shannon** 熵和降低 **HHI**

目标 **Q**(内容质量): 提升高风险内容质量门槛, 减少标题党和低俗内容

目标 **W**(用户福利): 用户获得符合其偏好的信息消费体验

section、2.2 基于用户福利的理论推演

Gentzkow Shapiro (2011) 在 ***Quarterly Journal of Economics*** 中提出了基于偏好满足度的用户福利测度。**Chu et al. (2023)** 在 ***Management Science*** 的最新研究进一步揭示,**AI** 个性化推荐对用户福利的影响高度依赖偏好异质性和算法透明度: 当用户偏好异质性高且算法透明度低时, 统一的推荐策略会导致显著的福利损失。我们借鉴这些框架, 构建算法治理的用户福利函数:

$$W = \sum_{i \in \{S,E,O\}} \theta_i \cdot U_i(A_i, Q_i)$$

其中,θ_i 为用户对内容类型 i 的偏好权重 (异质性参数),$U_i(A_i, Q_i)$ 为用户从类型 i 获得的效用, 取决于注意力份额 A_i 和内容质量 Q_i。关键洞察在于: 现有监管隐含假设所有用户偏好同质 (**uniform** θ_i), 但 **Gentzkow Shapiro (2011)** 的实证研究表明, 用户信息消费偏好存在显著异质性, 部分用户偏好严肃新闻, 部分用户偏好娱乐内容。

在用户偏好异质性条件下, 我们可以推演三条" 不可能三角" 路径:

路径 **1: D+Q → W** 不确定 (本研究实证验证)

监管同时推动信息多样性提升和内容质量改善:

多样性目标 **: Shannon** 熵 **+18.7**

质量目标 **:** 社会类信息差**-20.0**

但用户福利 **W** 取决于 θ_i 分布的异质性: **-** 若用户 **A** 偏好社会类 ($\theta_S^A = 0.5$), 监管后 A_S 从 **34.82-** 若用户 **B** 偏好娱乐类 ($\theta_E^B = 0.4$), 监管后 A_E 从 **16.47-** 总福利变化 $\Delta W = \Delta W^A + \Delta W^B$ 的符号不确定, 取决于偏好 θ_i 的分布

由于本研究无法观测用户层面的偏好参数 θ_i, 我们无法断言用户福利是否改善。但注意力零和约束保证了监管必然创造" 赢家" 和" 输家": 偏好社会类的用户受损, 偏好娱乐类的用户受益。

路径 **2: D+W → Q↓**(理论推演)

若监管同时追求多样性和用户福利 (允许用户自主选择):

用户可能选择低质量但高娱乐性的内容 (**Sunstein, 2001** 的" 信息茧房" 警告)

平台为追求用户停留时长而降低质量门槛

导致目标 **Q** 失效: 社会类质量难以提升

路径 **3: Q+W → D↓**(理论推演)

若监管同时追求质量和用户福利:

高质量内容供给有限 (社会类官媒内容池较小)

用户偏好多样, 平台需增加各类型供给以满足异质性需求

但质量约束限制了供给扩张, 导致多样性下降

目标 **D** 失效:**Shannon** 熵难以提升

section、2.3 打破" 不可能三角" 的理论路径

" 不可能三角" 并非绝对不可打破, 关键在于放松注意力零和约束或显性化用户偏好异质性。

路径一: 扩展注意力供给 (放松 $\sum_i A_i = \bar{A}$ 约束)

短期内注意力供给刚性 (**Wu Huberman, 2007**), 但长期可通过以下途径扩展: - 提升内容吸引力, 延长用户在平台停留时长 - 增加用户活跃度, 扩大日活跃用户规模 - 优化用户界面, 降低信息获取的认知成本

若 $\bar{A}$ 扩张, 则可在不牺牲某一类型份额的前提下同时提升多样性和质量

路径二: 显性化用户偏好 (引入 θ_i^{user})

当前监管假设 θ_i 同质, 但 **Gentzkow Shapiro (2011)** 证明异质性显著

若允许用户表达个性化偏好 θ_i^{user}, 平台可定制化推荐: - 偏好社会类的用户接收更多社会类内容 ($A_S^{user} > A_S^{avg}$) - 偏好娱乐类的用户接收更多娱乐类内容 ($A_E^{user} > A_E^{avg}$)

在全局多样性约束下, 个体可偏离平均分布, 实现个性化多样性

section、三、政策创新: 从供给侧监管到需求侧赋权

Constantinides et al. (2018) 在 *MIS Quarterly* 中强调, 平台治理需要平衡多方利益相关者 (监管者、平台、内容提供者、用户)。当前算法治理主要聚焦供给侧 (约束平台算法), 但本研究揭示的" 不可能三角" 表明, 必须转向需求侧 (赋权用户注意力配置), 才能在多样性、质量、用户福利之间取得更优平衡。

我们提出用户注意力主权 (**User Attention Sovereignty**) 框架, 核心主张:

注意力是用户的稀缺资源, 用户应拥有对自身注意力分配的决策权

算法应是用户的代理人, 而非平台或监管者的单向工具

多样性应由用户偏好聚合决定, 而非自上而下强制

基于这一理念, 我们提出四大政策支柱:

section、3.1 支柱一: 用户可控的算法透明化

理论基础: **Eslami et al. (2015, CHI)** 发现, 用户对算法的感知透明度影响其信任和满意度。**Vaccaro et al. (2024)** 在 *MIS Quarterly* 的最新研究进一步指出, 算法透明化需要" 可解释性"(**explainability**) 与" 可控性"(**controllability**) 并重: 单纯披露算法规则不足以建立用户信任, 用户需要实质性的控制权来调整算法行为。**Aral Dhillon (2024)** 在 *Information Systems Research* 中强调, 数字足迹揭示的用户真实偏好应成为算法优化的基础, 而非平台单方面推断的偏好。**Diakopoulos (2016, Digital Journalism)** 进一步指出, 算法问责机制需要技术透明与用户赋权相结合。

政策措施一: 个性化多样性配额

现状问题: 监管设定统一的多样性标准, 但用户偏好异质性被忽视

创新方案: 允许用户自定义社会类 **vs** 娱乐类的推荐比例 - 技术实现: 在算法中引入用户偏好参数 θ_i^{user}, 推荐结果为平台默认与用户偏好的加权:

$$A_i^{user} = \alpha \cdot A_i^{platform} + (1-\alpha) \cdot \theta_i^{user}, \quad \alpha \in [0,1]$$

- 用户可通过界面滑块调节 α:$\alpha = 1$ 为完全接受平台推荐,$\alpha = 0$ 为完全自主配置 - 类似 **YouTube** 的" 不感兴趣" 功能, 但更系统化和前置化

理论支撑:**Bakshy et al. (2015, Science)** 表明, 用户自主选择比算法过滤更能提升信息多样性。通过赋予用户配置权, 可在全局多样性约束下实现个性化多样性。

政策措施二: 注意力预算仪表盘

现状问题: 用户对自身注意力分配缺乏自我觉知 **(attention self-awareness)**

创新方案: 向用户展示其注意力分配结构 - 每周报告:" 您本周在社会类内容上投入 **X** 小时 (占比 **Y-** 类似 **Apple** 的" 屏幕使用时间", 但聚焦内容类别而非应用类别 - 提供历史趋势图, 帮助用户识别注意力分配的变化模式

理论支撑:**Thaler Sunstein (2008)** 的" 助推"**(nudge)** 理论表明, 简单的信息呈现可帮助用户做出符合长期利益的决策。注意力仪表盘通过提升自我觉知, 助推用户主动调整注意力分配。

section、3.2 支柱二: 偏好驱动的动态多样性标准

理论基础: **Gentzkow Shapiro (2011, QJE)** 证明用户信息消费偏好存在显著异质性。**Jiang et al. (2023)** 在 ***Management Science*** 的最新研究揭示了个性化推荐的" 暗面": 通过强化偏好反馈循环, 推荐系统会导致长期多样性下降, 形成" 信息茧房"**(filter bubble)**。**Anderson et al. (2024)** 在 ***Journal of Marketing*** 中进一步发现, 算法策展对消费者福利的影响呈倒 **U** 型, 过度个性化反而降低福利。这提示监管需要在" 个性化" 与" 多样性" 之间取得平衡。**Hosanagar et al. (2014, MS)** 警告, 统一的推荐策略可能加剧用户分化, 降低整体福利。

政策措施一: 分层多样性标准

全局多样性 (平台层面): 平台整体必须满足 **Shannon** 熵 阈值 (如 **0.8**), 保障信息生态的基本多样性

个体多样性 (用户层面): 单个用户的推荐可以偏离全局分布, 但需满足两个条件: - 定期推送" 反泡泡" 内容: 每周至少推送 **3-5** 条用户历史未接触类型的内容, 防止信息茧房固化 **(Pariser, 2011)** - 用户可选择" 探索"**vs**" 利用" 模式: - 探索模式 **(exploration)**: 算法优先推荐新颖内容, 提升多样性 - 利用模式 **(exploitation)**: 算法优先推荐偏好内容, 提升满意度 - 用户可根据情境切换模式 (如通勤时探索, 午休时利用)

政策措施二: 偏好学习与更新机制

现状问题: 平台通过用户行为 (点击、停留时长) 推断偏好, 但可能导致" 偏好固化"**(preference lock-in)**

创新方案: 引入偏好更新机制 - 定期随机推荐非偏好内容 (类似强化学习中的 ϵ**-greedy** 策略) - 用户可手动修正偏好估计 (如标注" 我不喜欢此类内容" 或" 我想多看此类内容") - 平台需向监管者披露偏好推断算法的准确率和更新频率

理论支撑:**Sen (1999)** 的" 能力方法"(**capability approach**) 强调, 福利不仅取决于实际消费 (**achieved functionings**), 更取决于选择自由 (**capability set**)。偏好学习机制扩展了用户的选择自由。

section、3.3 支柱三: 公共价值内容的正向激励

理论基础: **Eaton et al. (2015, MISQ)** 的边界资源理论表明, 平台可通过激励而非约束引导生态演化。**Karunakaran Orlikowski (2024)** 在 ***Organization Science*** 的最新研究强调, 平台问责需要构建多层次基础设施, 包括技术透明、经济激励、声誉机制的协同。**Cennamo et al. (2023)** 在 ***Academy of Management Review*** 中指出, 平台治理的核心挑战是开放性 (促进创新) 与控制性 (保障质量) 的平衡, 正向激励可以在保障质量的同时维持生态活力。**Rahman et al. (2024, AMA)** 的平台问责框架强调, 正向激励比负向约束更能实现长期可持续治理。

现状问题: 当前监管通过" 提升社会类审核成本" 抑制劣质内容, 但本文 **H2** 理论模型揭示的超模成本函数 ($\frac{\partial^2 C}{\partial Q \partial R} > 0$) 表明, 高风险内容的质量提升成本呈指数增长, 导致平台减少社会类供给以避免成本爆炸。这是典型的负向激励机制, 结果是" 按下葫芦浮起瓢"。

创新方案: 改变平台优化函数的收益项而非成本项:

$$\max \sum_i \left[\underbrace{R_i(A_i)}_{\text{商业收益}} + \underbrace{\beta \cdot PV_i(A_i, Q_i)}_{\text{公共价值补贴}} \right] - C_i(A_i, Q_i)$$

其中,$PV_i(A_i, Q_i)$ 为公共价值函数 (衡量内容的社会正外部性),β 为监管者设定的补贴系数。通过引入公共价值激励项, 平台在商业收益与公共价值之间取得平衡。

具体措施:

措施一: 流量扶持机制

对高质量社会类内容 (如深度报道、数据新闻、调查报道) 给予算法权重倾斜

技术实现: 在推荐排序中引入" 公共价值加权因子"w_{PV}, 使得优质社会类内容的排序分数为:

$$\mathbf{Score}_i = \mathbf{Score}_i^{base} \times (1 + w_{PV} \cdot Q_i), \quad w_{PV} \in [0.1, 0.3]$$

类似 **YouTube** 的" 新闻货架"(**news shelf**) 和" 权威来源"(**authoritative sources**) 标签

措施二: 经济激励机制

财政补贴: 按平台社会类优质内容占比发放补贴, 类似文化产业扶持政策

税收减免: 对达到公共价值评分阈值的平台给予企业所得税减免 (如减免 **1-3** 个百分点)

广告分成倾斜: 平台广告收入的一定比例 (如 **5-10**

措施三: 声誉激励机制

建立” 算法治理评级”, 定期公布平台公共价值得分 (类似 **ESG** 评级)

评级指标包括: 社会类优质内容占比、信息差指数、多样性指数、用户满意度

高评级平台获得政策优先支持 (如数据开放试点、国际合作优先)

理论支撑: **Parker Van Alstyne (2018, MS)** 在平台创新研究中指出, 平台控制权的优化配置需要激励相容 **(incentive compatibility)**。通过正向激励, 监管者将” 公共价值最大化” 内化为平台的自利行为, 实现监管目标与商业目标的一致性。

算法治理的合法性基础。**Kemper Kolkman (2023)** 在 ***Information and Organization*** 中指出, 算法治理的合法性来源于程序公正 **(procedural justice)** 而非单纯的结果优化。本研究发现的” 被动多样性” 现象 **(Shannon 熵 +18.7**

section、3.4 支柱四: 跨平台注意力账户与协同治理

理论基础: **Tiwana et al. (2010, ISR)** 的平台演化理论强调环境动态变化对平台策略的影响。**Jacobides et al. (2022, SMJ)** 的平台生态系统理论指出, 平台间竞争会通过用户多宿 **(multi-homing)** 放大网络外部性。

现状问题: 单平台治理可能引发用户跨平台迁移, 导致监管失效。本研究发现的” 水床效应”(社会类份额下降 $\rightarrow$ 娱乐类份额上升) 可能在平台间复制: 若微博监管严格, 用户可能转向抖音、小红书寻求娱乐内容, 形成” 监管套利”**(regulatory arbitrage)**。

创新方案一: 用户注意力账户

理念: 类似 **GDPR** 第 **20** 条的” 数据可携带权”**(data portability)**, 建立” 注意力偏好可携带权”

技术实现: - 用户在平台 **A** 设定的偏好参数 θ_i^{user}(社会类 **40-** 当用户转换至平台 **B** 时, 可导入该偏好配置, 平台 **B** 需尊重用户设定 - 类似 **OAuth** 的跨平台身份认证, 但聚焦偏好而非身份

政策支撑: 监管部门制定” 注意力偏好数据标准”, 要求主流平台支持偏好导入/导出功能

创新方案二: 跨平台协同监管机制

统一标准: 对社会类、娱乐类内容的质量标准和多样性要求实施跨平台统一 - 建立” 全国内容质量数据库”, 平台共享低质内容标识 (黑名单机制) - 防止劣质内容在平台间流转 (如微博删除的标题党内容转发至抖音)

数据共享: 在隐私保护前提下 **(privacy-preserving)**, 平台间共享聚合数据 - 采用联邦学习 **(federated learning)** 技术, 不共享原始数据但共享模型参数 - 识别低质内容的跨平台传播路径, 实施联合治理

联合执法: 监管部门协调多平台同步推进治理行动 - 避免” 按下葫芦浮起瓢”: 若仅监管微博, 用户可能转向抖音 - 同步监管形成” 监管合力”, 降低用户跨平台套利空间

理论支撑: **Rochet Tirole (2003)** 的双边市场理论指出, 平台竞争中的用户多宿会放大网络外部性。注意力账户可携带性降低用户转换成本, 促进平台间良性竞争 **(competition on quality)** 而非监管套利 **(competition on laxity)**。

section、四、研究局限与未来方向

section、4.1 数据与测度局限

用户偏好数据缺失。本研究仅观测到平台供给侧的注意力配置 (A_i), 无法直接测度用户需求侧的真实偏好 (θ_i^{user})。这限制了我们对用户福利变化的推断。**Gentzkow Shapiro (2011)** 的研究表明, 用户信息消费偏好存在显著异质性, 部分用户偏好社会类内容 (如政治新闻重度消费者), 部分用户偏好娱乐类内容 (如娱乐八卦爱好者)。当监管导致社会类份额下降 **3.16** 个百分点时, 前者福利受损, 后者福利可能改善 (娱乐类份额上升 **3.84** 个百分点)。但由于缺乏用户层面数据, 我们无法量化总福利变化 $\Delta W = \sum_i \theta_i \Delta U_i$ 的符号和大小。

跨平台数据缺乏。本研究聚焦微博单一平台, 无法验证用户是否因监管而迁移至抖音、小红书等平台。**Tiwana et al. (2010)** 的平台演化理论强调, 用户多宿 **(multi-homing)** 会改变平台间竞争格局。若微博监管严格但抖音监管宽松, 理性用户可能将注意力转移至抖音, 形成" 监管套利"。本研究发现的" 水床效应"(社会类 → 娱乐类) 可能在平台间复制为" 微博 → 抖音" 的用户迁移, 但缺乏跨平台数据验证。

长期效应未知。本研究观测窗口为政策实施后 **10** 个月 **(2024** 年 **11** 月至 **2025** 年 **10** 月), 仅能捕捉短期和中期效应。**Wu Huberman (2007)** 指出, 注意力分配存在适应性调整 **(adaptive reallocation)**, 生态系统可能在监管冲击后逐渐达到新均衡。本研究发现的趋势项 **(-0.0008/日,p<0.001)** 表明社会类份额持续下降, 但长期是否趋于稳定? 平台是否会开发新策略规避监管? 这些问题需要更长时间窗口的追踪研究。

section、4.2 机制识别局限

平台算法细节的黑箱问题。本研究无法观测平台如何具体调整推荐算法 (特征权重、排序规则、多样性约束的技术实现)。**Diakopoulos (2016)** 强调, 算法透明化是算法问责的前提。但在实际研究中, 平台算法的商业机密性质使得研究者难以获得技术细节。我们仅能通过" 输入-输出" 分析 (政策冲击 → 注意力配置变化) 推断平台行为, 无法打开" 算法黑箱" 验证具体机制。这限制了我们对 **H1**(去中心化如何实现) 和 **H3**(水床效应的算法路径) 的深入理解。

用户响应异质性未探索。本研究使用平台层面日度数据, 无法识别不同人口学特征用户的异质性响应。**Gentzkow Shapiro (2011)** 发现, 年龄、教育程度、政治倾向显著影响用户的信息消费偏好。理论上, 监管引发的注意力再分配 (社会类 ↓ 娱乐类 ↑) 可能对不同用户群体产生差异化影响: 高学历用户可能更依赖社会类内容获取公共信

息, 其福利损失更大; 低学历用户可能更偏好娱乐类内容, 其福利可能改善。但缺乏用户层面数据, 我们无法检验这一异质性假说。

因果识别的威慑效应问题。**ITS** 和 **DID** 方法假设政策断点清晰, 但算法治理存在”威慑效应”(**deterrence effect**): 平台可能在政策正式实施前就预期性调整算法。若威慑效应显著, 真实的政策效应可能被低估 (处理组在政策前已部分调整)。本研究通过安慰剂检验 (伪断点 **p=0.728**) 和平行趋势检验 (**p=0.421**) 部分缓解该问题, 但无法完全排除预期性调整的可能性。

section、4.3 未来研究方向

方向一: 用户偏好的实验测度

田野实验设计: 与平台合作, 向部分用户提供算法控制权 (如自定义社会类 **vs** 娱乐类比例), 观测其真实选择行为

研究问题: - 用户真实偏好 θ_i^{user} 的分布特征 (均值、方差、异质性来源) - 用户是否会主动调整注意力配置, 还是接受平台默认设置 (默认效应,**Thaler Sunstein, 2008**) - 赋予控制权后, 用户福利 W 是否改善

理论贡献: 为” 不可能三角” 提供用户偏好的微观基础, 验证” 需求侧赋权” 的有效性

方向二: 跨平台注意力溢出效应

数据收集: 同时收集微博、抖音、小红书等多平台的用户行为数据 (需与平台合作或使用爬虫 + 用户授权)

研究问题: - 监管是否导致用户跨平台迁移 (微博使用时长 ↓, 抖音使用时长 ↑) - 注意力溢出的方向和幅度 (社会类内容从微博转向哪个平台?) - 跨平台监管协调是否能降低溢出效应

理论贡献: 验证平台间竞争对算法治理效果的影响, 为跨平台协同治理提供实证依据

方向三: 用户福利的长期效应

追踪研究设计: 追踪监管后用户的信息素养、政治参与度、主观福利指标 (生活满意度、心理健康)

研究问题: - 娱乐类份额上升是否损害用户的信息获取能力和公民参与度 (**Sunstein, 2001** 的” 信息茧房” 假说) - 社会类份额下降是否削弱” 知情公民”(**informed citizen**) 的培育 - 多样性提升是否真正改善用户福利, 还是仅实现了” 被动多样性”

理论贡献: 从短期注意力配置转向长期用户福利, 评估算法治理的社会总福利效应

方向四: 算法治理的国际比较

比较案例: 对比中国 (清朗行动)、欧盟 (**DSA** 数字服务法案)、美国 (市场自律 + 反垄断) 的治理模式

研究问题: - ” 供给侧监管”(中国模式)vs” 需求侧赋权”(欧盟 **GDPR** 模式) 的效果差异 - 算法透明化要求 (**DSA** 第 **27** 条) 是否比算法约束 (清朗行动) 更有效 - 不同治理模式对平台创新、用户福利、内容多样性的长期影响

理论贡献: 为全球算法治理提供比较制度分析框架, 识别最优治理实践

section、五、结语

算法推荐系统作为数字时代的” 注意力分配器”, 深刻影响着信息环境、公共讨论和社会认知。本研究利用中国清朗行动这一准自然实验, 系统评估了算法治理的多维效应, 揭示了监管在实现多样性提升和质量改善的同时, 也引发了非预期的注意力再分配。我们提出的” 不可能三角” 理论框架表明, 在注意力零和约束下, 多样性、质量、用户福利三大目标之间存在内在张力, 单一维度的监管可能陷入” 按下葫芦浮起瓢” 的困境。

打破这一困境, 需要从” 供给侧监管” 转向” 需求侧赋权”, 将用户从算法治理的被动接受者转变为主动参与者。通过赋予用户对算法的控制权 (个性化多样性配额)、提升自我觉知 (注意力预算仪表盘)、建立正向激励 (公共价值补贴)、实现跨平台协同 (注意力账户可携带), 我们可以在保障信息多样性和内容质量的同时, 尊重用户偏好的异质性, 实现更高水平的用户福利。

算法治理是一个系统工程, 涉及监管者、平台、内容提供者和用户多方博弈。本研究为理解这一复杂系统提供了新的理论视角和实证证据, 也为完善中国乃至全球的算法治理体系提供了政策启示。未来研究需要进一步打开” 算法黑箱”, 测度用户真实偏好, 追踪长期福利效应, 并通过国际比较识别最优治理实践。唯有如此, 我们才能在数字时代构建一个既充满活力又负责任的信息生态系统。

References

Boudreau, K. J. and Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. *Platforms, Markets and Innovation*, pages 163–191.

Gorwa, R., Binns, R., and Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1).

Gritsenko, D. and Wood, M. (2022). Algorithmic governance: A modes of governance approach. *Regulation & Governance*, 16(1):45–62.

Huber, T. L., Fischer, T. A., Dibbern, J., and Hirschheim, R. (2017). A process model of complementarity and substitution of contractual and relational governance in is outsourcing. *Journal of Management Information Systems*, 34(1):81–114.

Jacobides, M. G., Cennamo, C., and Gawer, A. (2022). Towards a theory of ecosystems. *Strategic Management Journal*, 43(3):543–572.

Parker, G. G. and Van Alstyne, M. W. (2018). Innovation, openness, and platform control. *Management Science*, 64(7):3015–3032.

Tiwana, A., Konsynski, B., and Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4):675–687.

Ulbricht, L. and Yeung, K. (2022). Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation & Governance*, 16(2):466–492.

Wang, L., Chen, J., and Gu, X. (2025). Platform governance and content quality: Evidence from social media. *Information Systems Research*. Forthcoming.